



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Are trait-outcome associations caused by scales or particular items? Example analysis of personality facets and BMI.

Citation for published version:

Vainik, U, Möttus, R, Allik, J, Esko, T & Realo, A 2015, 'Are trait-outcome associations caused by scales or particular items? Example analysis of personality facets and BMI.', *European Journal of Personality*, vol. 29, no. 6, pp. 622-634. <https://doi.org/10.1002/per.2009>

Digital Object Identifier (DOI):

[10.1002/per.2009](https://doi.org/10.1002/per.2009)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

European Journal of Personality

Publisher Rights Statement:

This is the peer reviewed version of the following article: Vainik, U., Möttus, R., Allik, J., Esko, T., & Realo, A. (2015). Are trait-outcome associations caused by scales or particular items? Example analysis of personality facets and BMI. *European Journal of Personality*, which has been published in final form at <http://dx.doi.org/10.1002/per.2009>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Are trait-outcome associations caused by scales or particular items?

Example analysis of personality facets and BMI.

Uku Vainik* ^a; René Mõttus ^{a, b}; Jüri Allik ^{a, c}; Tõnu Esko ^d, Anu Realo ^a

^a Institute of Psychology, University of Tartu, Näituse 2, Tartu, 50409, Estonia

^b Department of Psychology, Centre for Cognitive Ageing and Cognitive Epidemiology,
University of Edinburgh, 7 George Square, Edinburgh, Scotland, EH8 9JZ

^c Estonian Academy of Sciences, Kohtu 6, Tallinn, 10130, Estonia

^d Estonian Genome Centre of University of Tartu, Estonia

* Corresponding author, uku.vainik@gmail.com , +372 737 6549

Short title (70 chars)

Are trait-outcome associations caused by scales or single items?

Abstract 200 words

In personality research, trait-outcome associations are often studied by correlating scale sum-scores with an outcome. For example, an association between the NEO Impulsiveness scale and body mass index (BMI) is often interpreted to pertain to underlying trait Impulsiveness. We propose that this expectation can be corroborated by testing for Spearman's theorem of indifference of indicator. Namely, an underlying trait-outcome association should not depend on the specific items (i.e., indicators) used to measure the trait. To test this theorem, we outline an indicator exclusion procedure and demonstrate its viability using a simulation design. We then apply this procedure to test personality-BMI associations for indifference of indicator in a large population-based sample of adult Estonians (N=2581) using self- and informant ratings obtained with the NEO Personality Inventory-3. Our results show that the N5: Impulsiveness-BMI association mostly depends on two eating-related items, suggesting that the trait associated with BMI may be narrower than the trait the N5: Impulsiveness scale is supposed to measure. Associations between BMI, E3: Assertiveness and C2: Order seem to pertain to the trait. In sum, testing for indifference of indicator provides a potentially useful method to clarify trait-outcome relationships. R scripts are provided that implement the indicator exclusion procedure.

6 keywords

Indifference of indicator, Spearman, reflective, Impulsiveness, single item, sum score

1 Introduction

A common method to understand human behaviour is relating various outcomes of interest to personality traits (Ozer & Benet-Martínez, 2006; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). In such analyses, a trait is often operationalized as sum-scores of a personality scale consisting of individual items. These individual items of a scale have been chosen to best represent and quantify an ostensibly underlying latent trait – the true personality attribute. Correlations of the scale sum-scores with outcomes of interest are then interpreted as pertaining to the underlying trait. However, we argue that such interpretations may sometimes merit more careful consideration, particularly if traits are indeed conceptualized as latent variables.

In latent variable framework, the true variance of a personality trait is manifested in the common variance of scale's items (further described below). If so, to the extent that a correlation between a scale's sum-score and an outcome reflects the correlation of the outcome with an underlying trait, the association should be independent of how the underlying trait is measured. In particular, the association should not depend on which items happen to be sampled for the measurement of the trait. Otherwise the association would not pertain to the underlying trait, but to the specific characteristics reflected in the particular items driving the observed correlation. This may sound like a trivial assumption, but it is often neglected in practice.

For example, the N5: Impulsiveness facet of the NEO Personality Inventory (NEO-PI-R) has been related to body mass index (BMI) in several studies (Sutin, Costa, et al., 2013; Sutin et al., 2015; Sutin, Ferrucci, Zonderman, & Terracciano, 2011; Terracciano et al., 2009).

However, not all N5: Impulsiveness items relate to BMI. Terracciano et al (2009) specify that “the Impulsiveness scale effect [on BMI] was due to two items specific to the eating domain

(“When I am having my favorite food, I tend to eat too much” and “I sometimes eat myself sick””) (Terracciano et al., 2009, p. 685). Although this may seem to provide little evidence for trait Impulsiveness being associated with BMI, follow-up studies have still concluded that Impulsiveness relates to BMI without testing the crucial role of the two above-mentioned items behind this association (Sutin, Costa, et al., 2013; Sutin et al., 2011, 2015; see Iacovino, Powers, & Oltmanns, 2014 for a notable exception).

It should be noted that the same logic has already been applied in the context of the Big Five domains and their facets. In particular, there is ample evidence that different facets of the same Big Five domain scales correlate with outcomes in different ways (e.g., Costa & McCrae, 1995; Judge, Rodell, Klinger, Simon, & Crawford, 2013). Therefore, facets reflecting the same trait should relate to outcomes in similar ways *if* the associations pertain to the underlying domain-level trait. If they do not, then the associations pertain to facet(s) and not to the domain-level trait. This logic is commonly accepted as can be seen in Sutin et al (2011): although they demonstrate that BMI relates to trait Neuroticism, they appropriately interpret the effect being driven by the N5: Impulsiveness facet. We propose that the same logic should be extended to indicators of single traits or facets, that is to single items.

There is increasing evidence that the personality trait hierarchy may be meaningfully extended below facets. A recent analysis of cross-rater agreement across broad domains, facets, and single items concluded that there was a significant amount of descriptive variance left in NEO-PI-3 items that raters agreed on, once the effects of broad domains and facets had been removed (R. Mõttus, McCrae, Allik, & Realo, 2014). Similarly, another paper attempted building unidimensional CFA models of NEO-PI-3 facets; to achieve reasonable fit indices, almost all facets required allowing additional residual correlations between items (see S2 in René Mõttus et al., 2015), demonstrating additional co-variance—and thereby 'signal'—below the facets. McCrae (2014) suggested that trait-like characteristics below the level of

facets should be called nuances. As these nuances have not been documented yet, we reserve our analysis to items. Similarly to nuances, single items might reflect different small trait-like features.

The benefit of understanding what causes a scale-outcome association is schematically shown in Figure 1. When there is a domain-level association with an outcome, then all the sub-indicators should be correlated with the outcome as well (Figure 1A). When the association pertains to the facet instead of the domain (Figure 1B), then there might be a relatively weaker domain-level correlation. However, it is more precise to claim that the facet is causing the trait-outcome association, not the broad domain. Similarly, if there is a strong true association between single item (or nuance) and an outcome, a diluted effect might be visible at facet or even at domain level. However, it would be more precise to attribute the association to the nuance.

Figure 1 about here

1.1 Indifference of indicator: Theoretical background.

The logic outlined above originates from Spearman's (1927) theorem called *indifference of indicator* (ION¹). He introduced ION in the context of measuring general intelligence (*g*): “for the purpose of indicating the amount of *g* possessed by a person, any test will do just as well as any other, provided only that its correlation with *g* is equally high” (Spearman, 1927, p. 197). In intelligence research, this theorem has enjoyed notable popularity (reviewed in Horn & McArdle, 2007) and has received strong empirical support (Johnson, Bouchard Jr.,

¹ Indifference Of iNdicator

Krueger, McGue, & Gottesman, 2004; Johnson, Nijenhuis, & Bouchard, 2008). We propose that it could equally be applied in personality research. In particular, we articulate some conditions which must be fulfilled for a trait-outcome association to be indifferent of its indicators. We also propose a statistical procedure that tests whether a trait-outcome association is indicator indifferent. We discuss ION in the context of single items being indicators for a particular trait, but the same logic also applies, in principle, when facets are considered indicators of a broader latent domain.

Here, we take the assumption that personality traits are of reflective type, whereby the trait is a latent common cause of its indicators (e.g., items) (Bollen & Lennox, 1991; Borsboom, Mellenbergh, & van Heerden, 2003). Although we acknowledge that there may be other views as to the nature of personality traits and embrace them, assuming traits to be reflective seems to be the most common trait interpretation (Borsboom, 2006; DeYoung, 2014; McCrae, 2014). According to this interpretation, the true variance of a trait is manifested in the common variance of its indicators. The common variance can be explicitly modelled, as in case of latent trait modelling, or assumed to emerge via aggregation, as in case of classical test theory applications. The indicators are “noisy” and the degree to which they reflect the underlying latent trait can be expressed in terms of factor loadings or item-total correlations.

Consequently, if an outcome correlates with a trait, none of the trait’s indicators (i.e., items) should correlate with the outcome more than the aggregate trait score. This is because the associations between single indicators and the outcome are indirect, mediated by the underlying trait. Single indicator-outcome correlations should then equal the products of the trait-outcome correlation and the factor loadings of the respective indicators. Therefore, the single indicator-outcome correlations should be proportional to their factor loadings, *if* it is only the latent trait that correlates with the outcome. However, this might not always be the

case, as seen in the N5: Impulsiveness- BMI association addressed above. It seems that personality scales have not always been explicitly designed to be unidimensional (c.f., Gerbing & Anderson, 1988; Gorsuch, 1997). Therefore, the items in a scale might have other sources of covariance than the trait they have been intended to measure. When such a scale is related to the outcome, it might not be the main trait, but these other sources of covariance that cause the correlation between a scale and an outcome. In the case of N5: Impulsiveness, it seems that only eating-related items and not generic impulsivity items are associated with BMI. The question then is this: How can we test if an outcome relates to the trait itself or only to a certain subset of its indicators?

1.2 Methods for detecting indifference of indicator

A natural test of ION would be to measure a purported trait with multiple scales that sample different indicators of the trait and have roughly similar reliabilities or determinacies: if ION holds, their correlations with the outcome should be similar (René Möttus, Luciano, Starr, Pollard, & Deary, 2013). However, researchers can rarely afford to measure traits with multiple scales.

One may then estimate the presence of ION by correlating the outcome of interest to single items. This is a very straightforward procedure, but appears rarely employed in practice. If most items of a scale correlate with the outcome and the correlations are in the same direction, this may be consistent with the presence of ION; preferably, no single item-outcome correlation should exceed the scale-outcome correlation, as was discussed above. However, this method does not provide a formal test for deciding, whether some items violate the principle of ION. For example, some variability across items in their outcome-correlations is expected due to differential factor loadings or measurement noise.

To obtain a more formal test of ION, one could employ the “method of correlated vectors” (Jensen, 1998). According to this method, there should be a positive correlation between the factor loadings and outcome-correlations of trait indicators. Indicators deviating from the regression line can be considered to violate ION. However, this method has been criticized on several grounds. For example, it appears to be severely biased in favour of observed associations pertaining to the underlying trait (Lubke, Dolan, & Kelderman, 2001), but also sensitive to the selection of trait indicators (Ashton & Lee, 2005).

Therefore, a procedure devoted to testing ION is needed. We propose a simple *indicator exclusion* approach that provides an opportunity to statistically test, whether a scale-outcome relationship meets the ION assumption and, if it does not, which items are causing the lack of ION. Namely, stemming from the principles of ION, it should not matter which items happen to be used to capture the trait. Therefore, one can systematically exclude any single item from a scale and observe the resulting changes in trait-outcome correlations. The correlations between original trait-outcome and reduced-trait outcome trait should not differ much from each other if it is the underlying trait that is related to the outcome. However, these correlations are likely to vary more if only certain items of a scale relate to the outcome. For instance, in the example of N5: Impulsiveness, if it is trait Impulsiveness that relates to BMI, the scale-outcome correlation should not change much, when the eating-related items are removed from the scale. However, if such removal causes the scale-outcome correlation to change or vanish altogether, this implies that the scale’s association with BMI is to some extent or entirely specific to certain items.

The detailed procedure is as follows. For each item, correlation between the scale’s sum-scores and outcome is calculated such that the particular item is excluded from the sum-

scores². Each of the obtained correlations will then be compared to the original scale-outcome correlation (sum-score of all items). This comparison can be conducted with William's test for two dependent correlations that share one variable (Steiger, 1980). William's test characterises difference between correlations with a p -value – a small p -value indicates that the tested difference between correlations is unlikely to have happened by chance, and could be considered a real difference. Thus, each item will receive a p -value characterising the “significance” of difference between correlations – here called “*significance of indicator exclusion*” (SONE³).

The items are then ordered according to the SONE. If the lowest SONE is above a certain p criterion (defined below), the trait-outcome relationship can be considered indicator indifferent. Conversely, if there is an item with a SONE below a conventional criterion, partial or complete lack of ION is likely due to that particular item. Therefore, the item with the lowest SONE should be excluded from the scale to establish an indicator indifferent trait-outcome relationship. The whole procedure is then repeated on the remaining set of items to determine if any additional items are compromising ION. The procedure stops if no more items obtain a SONE below the criterion. As a result, the list of excluded items represents items that have a correlation with the outcome *independent* of the scale's ostensible underlying trait.

As always, there is a question of an optimal criterion p -value. The significance criterion has to be a trade-off between Type I and Type II errors, given sample size. In smaller samples, the power to detect correlation difference is smaller, and therefore a more lenient criterion is needed (which obviously increases Type I error rate). In large samples, the procedure is likely

² The principle is similar to the routine of “Cronbach alpha if item excluded”.

³ Significance Of iNdicator Exclusion

to flag very small changes in correlation as significant, even if it does signal only trivial lack of ION. To establish an optimal significance criterion for a given sample size, a simulation was conducted. The simulation modelled scale-outcome associations that either had or did not have ION. An optimal cut-off criterion was selected that minimised the likelihood of excluding items if the association was indicator indifferent, and maximised the likelihood of excluding items if the association was lacking ION.

1.3 Aims of the paper

The current paper seeks to outline the method and the usefulness of testing trait-outcome associations for ION. Study 1 used simulations to demonstrate and test the viability of the proposed indicator exclusion method, and establish appropriate p -value criteria. In Study 2, the procedure was applied to empirical data by means of replicating BMI-personality trait associations (Terracciano et al., 2009) in a large population-based sample of adult Estonians. Scales that were significantly correlated with BMI were tested for ION. The BMI-personality relationship was chosen as some papers have explicitly tested if Impulsiveness -- BMI association depends on a few items (Iacovino et al., 2014; Terracciano et al., 2009) of the N5: Impulsiveness scale whereas other papers have not (Sutin, Costa, et al., 2013; Sutin et al., 2011, 2015). The BMI-personality dataset includes both self- and informant reports. Given that informant-ratings are considered as reliable as self-ratings of personality (Kolar, Funder, & Colvin, 1996), we assume that any robust association will replicate across both rating types. This could be taken as a case of constructive replication (Lykken, 1968).

2 Study 1: Simulation study

The goal of the simulation study was to show that ION could be tested with the indicator exclusion approach. The procedure was tested on simulated scales and outcomes, whose relations were or were not indifferent of indicator. We expected that the procedure would

correctly highlight items that caused lack of ION between scales and outcomes. Secondary goal of the simulation study was to provide relevant significance criteria for study 2.

2.1 *Simulation methods*

A trait scale and outcome were simulated according to two sets of scenarios (Figure 2). First, consistently with ION, scores of a single normally distributed trait (t) were simulated ($N = 1000, \mu = 0, \sigma = 1$). This trait was allowed to contribute to an outcome (o) with a standardized regression weight β , such that $o = \beta * t + \varepsilon_t$, where ε_t was a normally distributed random error term ($\mu = 0, \sigma = 1$). We tested three different β values 0.15, 0.25, and 0.35. Main focus is on $\beta = 0.25$ as such seems to be the effect size of trait N5: Impulsiveness- BMI association in Study 2. The simulated trait was manifested in a scale consisting of 8 items. The items had factor loadings λ ranging from 0.4 to 0.7; arguably, these may be seen as rather desirable parameter values (e.g., Ford, MacCallum, & Tait, 1986). We chose to have 8 items per scale, similar to the NEO-PI scales, so that the results of Study 1 could be applied in Study 2. This scale is referred to as “8+0”, indicating that scale does not have any items referring to a secondary trait.

A second set of scenarios reflected lack of ION due to the lack of unidimensionality in the scale. Instead of a single underlying trait, a composite scale consisting of two underlying correlated traits (T_1 and T_2) was created. Here, T_1 represented the main trait that the scale was purported to measure, whereas T_2 represented a trait related to but distinct from T_1 . Crucially, only T_2 was allowed to contribute to the outcome o , which mimics the situation where only a subset of a scale drives the correlation with an outcome. We tested three association strengths: $\beta = 0.15, 0.25$ or 0.35 . Otherwise, the same procedure as above was followed. The correlation between T_1 and T_2 was set at 0.3, reflecting rather typical inter-facet correlation of personality questionnaires (Ostendorf & Angleitner, 1994). Three versions of an 8-item scale lacking ION were simulated, in which 7 to 5 items were manifestations of T_1 , and, respectively, 3 to 1

items were manifestations of T_2 . Otherwise the same item generation procedure as above was used. These three scales lacking ION are referred to as “7+1”, “6+2”, and “5+3”.

Figure 2 about here

All scenarios were simulated 10,000 times. Within a simulation, scale-outcome associations were analysed with the indicator exclusion procedure. In scale with ION (8+0), no item was excluded. However, we did record the lowest SONE value that was necessary for obtaining an optimal p - value (see below).

A similar procedure was conducted with scales without ION (7+1, 6+2, 5+3). Besides calculating the lowest SONE value, we also designed the procedure to remove the item with the lowest SONE and then repeat itself, until no items were supposedly left in T_2 . For instance, in the case of 7+1 we first calculated the lowest SONE, then excluded the respective item, and calculated the lowest SONE again based on 7 items (referred to as “7+0”). To estimate the accuracy of the procedure, we inspected whether the item excluded did in fact belong to T_2 . A similar procedure was conducted on 6+2, where one item was excluded first (6+1), and another excluded thereafter, leaving a scenario where no more items should be removed (6+0).

As stated in section 1.2, the optimal p-criterion should be low enough to minimise the likelihood of excluding items when ION in fact exists. Therefore, p-criterion has to be lower than the lowest SONE value in a scenario where no items should be removed from the scale (8+0, 7+0, 6+0, 5+0). At the same time, the criterion should be high enough to maximise the likelihood of excluding items if the association is lacking ION. Therefore, the p-criterion has

to be higher than the lowest SONE value in scenarios lacking ION (7+1, 6+2, 5+3, but also 6+1, 5+2, and 5+1).

For all SONE mean values, we also depicted 95% confidence intervals. All simulations were conducted on sample sizes 100, 250, 500, 750, 1000, 2500, and 5000 and with three different β values. Analyses were conducted in R environment (R Core Team, 2013), occasionally relying on ‘psych’ package (Revelle, 2014).

2.2 *Results and Discussion*

Table 1 shows how the item exclusion’s accuracy depends on both the effect size in question and sample size. For larger effects ($\beta = 0.35$), violations of ION can be reasonably accurately (95%) detected in samples of 500 or more, whereas for $\beta = 0.15$, one would need a sample size of 2,500 or more to achieve a comparable level of accuracy. Although the item exclusion procedure requires relatively large samples if expected effect sizes are small, it may be a promising method to detect items causing lack of ION.

Table 1 about here

However, in real data we do not know a priori how many items need to be removed. Therefore, p -criteria are needed for one to be able to decide how many and which items to exclude. To define the maximum and minimum criteria, the average SONE values were calculated across simulations for all sample sizes. The results for $\beta = 0.25$ are depicted in Figure 3. It can be seen that mean SONE was greater for scales with ION (solid lines in Figure 3 B-D) than for scales without ION (dashed lines in Figure 3 B-D) for all sample sizes. This suggests that an optimal p -value can be found that lies between those two extremes.

Further, once the samples became 1,000 or more, the confidence intervals for the scales that lack ION (dashed lines in Figure 3 B-D) became smaller, suggesting that lack of ION could be more clearly detected in larger samples. Similar figures $\beta = 0.15$ and $\beta = 0.35$ can be seen in supplementary materials (Figures S1 and S2).

Figure 3 about here

To define the maximum values of the p-criteria, we first identified the lowest mean SONE across scenarios where no items had to be removed, which in case of $\beta=0.25$ corresponded to the lowest solid line across Figure 3 (i.e., Figure 3A). Thereafter, to define the minimum values for the p-criteria, we identified the highest SONE values when a trait was lacking ION, which corresponded to the highest dashed line across Figure 3 (i.e., 1+5 scenario in Figure 3D). These two scenarios have been plotted again in Figure 4.

Figure 4 about here

Apparently, in smaller samples, it was impossible to find an optimal p -value, as minimum criteria were larger than maximum criteria (Figure 4). In case of $\beta=0.25$, 750 was the minimum sample size from which onwards it was possible to obtain p -values (Figure 4). A likely reason is that with smaller sample sizes, it is very hard to reliably exclude the correct item in a 5+1 scenario – the uncertainty of this scenario is also illustrated by relatively wide confidence intervals in Figure 3D and with lowest percent of correctly excluded items in

Table 1 . Where it was possible to find an optimal p -value, this value was determined by the geometric mean between maximum and minimum. Optimal p -values for different effect sizes and scenarios are provided in Table 2. Here is an example with a sample size of 2500 and $\beta=0.25$: If removal of an item causes the scale-outcome correlation to change with a significance below 0.0037, then this item is causing violation in ION and has a separate relationship with the outcome.

Table 2 about here

3 Study 2: Personality – BMI relationships

Next, we sought to apply the tools outlined in the simulation analysis to study the presence or absence of ION in personality scale-BMI associations. We first screened for personality traits that related to BMI in both self-and informant report. This also enabled us to verify whether ION testing would replicate.

3.1 Methods

3.1.1 Participants

Participants were drawn from the Estonian Genome Center, University of Tartu (EGCUT). The EGCUT was launched as an initiative of the Estonian Government in 2001 to create a database of health, genealogical, and genome data representing 5% of the Estonian population (Leitsalu et al., 2014). EGCUT participants were randomly selected from individuals visiting general practitioners (GPs) and hospitals, recruited by GPs and hospital physicians. All participants gave informed consent. In addition to donating blood samples and answering a medical questionnaire, participants were asked to complete the self-report version of a

comprehensive personality test and find a knowledgeable informant who could complete the same questionnaire about them.

In total, the sample used in the present study included 2,581 people (of whom 1,398 were women) with a mean age of 44.0 years ($SD = 17.3$, ranging from 18 to 90 years) and a mean BMI of 26.08 ($SD = 4.9$, ranging from 15.9 to 54.1). Participants' weight and height were measured when they were recruited. Percent overweight ($BMI > 25$) in this sample is 52.9%, which matches a survey-based prevalence estimation of 49%, based on 5,000 Estonians (Tekkel & Veideman, 2013). However, another estimate that objectively weighed 495 participants representing the population suggested that prevalence of overweight status might reach 67% (Eglit, Ringmets, & Lember, 2013). Of the 2,581 participants, 8.2% people had elementary education, 24.5% had secondary school education, 27.9% had secondary specialised education, and 39.4% had a higher education degree. Of the informants, 52% were spouses or partners, 15% friends, 15% parents, 6% children or grandchildren, 6% siblings, 3% acquaintances, and 3% other relatives. The informants were 72% female and the mean age of informants was 42.4 years ($SD = 16.1$, ranging from 11 to 89 years). Overall, 5.6% of informants had elementary education, 25% had secondary education, 27.1% had secondary specialised education, and 42.2% had a higher educational degree.

3.1.2 Measures

The NEO Personality Inventory-3 (NEO PI-3; McCrae & Costa, 2010) is a slightly modified version of the NEO PI-R questionnaire (Costa & McCrae, 1992) that was translated into Estonian by Kallasmaa and colleagues (Kallasmaa, Allik, Realo, & McCrae, 2000). Like the original NEO PI-R, the NEO PI-3 has 240 items which measure 30 personality traits grouped into the FFM domains. The NEO PI-3 has excellent psychometric properties in a wide range of countries, including Estonia (De Fruyt, De Bolle, McCrae, Terracciano, & Costa, 2009). Participants themselves completed the self-report form of the NEO PI-3. Informants

completed the observer report form. In line with typical findings (Connolly, Kavanagh, & Viswesvaran, 2007), the correlations between the respective scale scores based on self- and informant reports were 0.53, 0.66, 0.61, 0.47, and 0.53 for Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness, respectively, and ranged from 0.39 to 0.62 (median = 0.46) for the 30 facet scales. For single items, the self-informant report correlations ranged from 0.13 to 0.56 (median = 0.30). All reported correlations were significant at $p < .01$.

3.1.3 Analytic strategy

BMI was log-transformed due to its skewed distribution and residualised for the effects of age, gender, and education, as these variables might have confounded personality-obesity relationships (Armon, Melamed, Shirom, Shapira, & Berliner, 2013; Brummett et al., 2006; Ogden et al., 2006; Rolls, Fedoroff, & Guthrie, 1991; Sutin, Costa, et al., 2013; Tekkel & Veideman, 2013). First, a correlation analysis of BMI and NEO-PI-3 domain and facet scales was performed as a replication of earlier studies (Sutin et al., 2011; Terracciano et al., 2009). Only scales significantly ($p < 0.01$) correlating with BMI in both self- and informant reports were taken further for the ION testing procedure. **Error! Bookmark not defined.** For ION testing, the optimal p criterion was chosen based on the above-described simulation results, considering our sample size and assumed trait-outcome association (β), which was set close to the highest single item-outcome correlation. We preferred single item correlations because if a scale violated ION, then the sum-score outcome correlation could be misleadingly low (see Figure 1). In the indicator exclusion procedure, indicators causing violation in ION were excluded iteratively until no SONE value was below the criteria. The results were plotted in a single item-outcome correlation plot, with excluded items highlighted (Figure 5, “single indicator-outcome correlation”). To further demonstrate that the excluded items might have their own relationship with the BMI, the scale-BMI relationship was graphed with and

without the excluded items (Figure 5, “Scale-outcome correlation). Some scale-BMI correlations were too small to be properly tested for ION; in these cases, only the single item-outcome correlations were plotted for preliminary assessment of ION (Figure 6). Barplots were plotted using ‘gplots’ package (Warnes et al., 2014)

3.2 *Results*

3.2.1 Correlations between BMI and personality scales

Table 3 lists the generally weak correlation coefficients between BMI and personality scales in both self and informant ratings. Three facets correlated with BMI in both self- and informant report and were thus selected for further analysis: BMI related positively to N5: Impulsiveness and E3: Assertiveness, and negatively to C2: Order.

Table 3 about here

3.2.2 Testing scale- BMI associations for ION

The optimal p criterion for N5: Impulsiveness was 0.0037, as some items approached a correlation of 0.25 with BMI (Table 2, Figure5). The N5: Impulsiveness failed to meet ION (Figure 5). In particular, indicator exclusion procedure suggested that scale’s relationship with BMI depended on two eating-related items⁴ (#111 – “I tend to eat too much of my favourite food” and #171 – “Sometimes I am not able to control my appetite”), as well as a more general impulse control item (# 51 – It is hard for me to control my impulses).

4

All items mentioned in the text from here on are back-translated versions of the NEO-PI-3 items from Estonian to English.

Figure 5 about here

The effects of other scales were too small for properly testing for ION. Simulations with $\beta = 0.10$ and $n=2500$ had revealed that percentage of correctly excluded items could range from 60% to 85% and that optimal p criterion could not be determined. However, for a preliminary assessment we plotted the single item-outcome correlations. Figure 6 suggests that, in contrast to N5: Impulsiveness, the associations of E3: Assertiveness and C2: Order scores with BMI were more likely to pertain to the core traits of the scales as over half of the items related to the outcome in a roughly equal level. At the same time, there were small inconsistencies as some items had very low correlation – future studies with higher power can formally test these associations for ION.

Figure 6 about here

3.3 Discussion

Testing personality-BMI relationships revealed that the well-documented correlation between BMI and Impulsiveness depended on a subset of items, mostly those relating to eating. The associations therefore fails to meet the assumption of ION. The effects of E3: Assertiveness and C2: Order were more likely to be indicator indifferent, but we were unable to formally test these associations due to small effect sizes. All these effects manifested in both self- and informant reports, supporting the robustness of the findings.

There could be two explanations for the lack of ION in case of N5: Impulsiveness-BMI association. First, the two eating-related items could reflect eating-related impulsivity, a

construct that has been suggested to be more BMI-relevant than domain-general impulse control across various measures (Houben, Nederkoorn, & Jansen, 2013; Rasmussen, Lawyer, & Reilly, 2010; Tsukayama, Duckworth, & Kim, 2012; see also Vainik, Dagher, Dubé, & Fellows, 2013). For instance, the three BMI-related items from Impulsiveness relate strongly to other eating-related scales (Vainik, Neseliler, Konstabel, Fellows, & Dagher, 2015).

Another interpretation could be that items asking about overeating are logically so close to BMI as to make any meaningful conclusion difficult. Namely, people might observe first that they are overweight and then conclude that they are unable to control themselves. For instance, N5: Impulsiveness is known to change in parallel with weight status (Sutin, Costa, et al., 2013). Whichever stance is taken, either of these interpretations is more precise than claiming that trait Impulsiveness is the underlying attribute relating to BMI. Based on these results, there would be no point saying it is impulsiveness as some sort of underlying trait that correlates with BMI and this is an important finding in its own right.

The sum-score of N5: Impulsiveness has been related to several other interesting outcomes, including BMI change, eating behaviours and disorders, leptin levels, white blood cell counts, drug and alcohol consumption, gambling, and brain activity, such as dopamine secretion and reward responsiveness (Bagby et al., 2007; Kristina Elfhag & Morey, 2008; Jen, Saunders, Ornstein, Kamali, & McInnis, 2013; Oswald et al., 2007; Ruiz, Pincus, & Dickinson, 2003; Sutin, Costa, et al., 2013; Sutin et al., 2012, 2011; Sutin, Zonderman, et al., 2013; Sutin, Evans, & Zonderman, 2013; Villafuerte et al., 2012). It would be interesting to reanalyse these effects for ION to understand if these outcomes pertain to trait Impulsiveness as the underlying attribute, or something more specific (c.f. Iacovino et al., 2014; Terracciano et al., 2009).

The effects of E3: Assertiveness and C2: Order on BMI are very small, but they have been repeatedly found across several studies (Sutin et al., 2011; Terracciano et al., 2009). The

effect of Order suggests that more organized persons have lower risk for obesity. A potential mechanism could be consistency in eating patterns – having similar meals across eating episodes has been shown to relate positively to successful weight maintenance and other health indices (Gorin, Phelan, Wing, & Hill, 2003; Pachucki, 2012; Wing & Phelan, 2005; see Vainik, Dubé, Lu, & Fellows, 2015 for further discussion). The effect of assertiveness seems to be instrument specific; studies with other personality instruments suggest instead that lack of assertiveness relates to maladaptive eating behaviours (K. Elfhag, 2005; K. Elfhag & Erlanson-Albertsson, 2006). This, obviously, points to potential lack of ION, which could be tested by linking items from multiple assertiveness scales to BMI scale.

4 General Discussion

The current paper has outlined a procedure to test if a trait-outcome association pertains to the whole scale or to its particular items. Guided by Spearman's (1927) theorem of indifference of indicator we suggest that all indicators of an underlying trait should similarly relate to the outcome. To apply this theorem in a personality context, we designed an indicator exclusion procedure that tests whether exclusion of an item significantly influences scale-outcome correlation.

Testing for ION is likely to be widely applicable, given that most personality trait-outcome research has so far been exclusively based on linking outcomes with aggregate scores. The point we want to emphasize is that the test entails win-win situations. In cases where the observable trait-outcome associations appear indicator-indifferent, the findings may appear even more robust. Otherwise, testing for ION may result in a more detailed description of the personality characteristics relating to focal outcomes. Therefore, while the sum-score approach can be used, supplementing it with ION testing provides greater confidence in or a better understanding of the observed effects.

Are some personality scales more likely to lack ION with outcomes? There is no clear answer, as various personality tests have been built using very different psychometric standards (Borsboom, 2006). Some scales have been constructed using rigorous standards (Gerbing & Anderson, 1988) which has tended to result in unidimensional scales.

Unidimensional scales may be more likely to have indicator indifferent relationship with an outcome, since the scales have been designed to reflect a single trait. In contrast, a scale-outcome association is more likely to lack ION if a scale has been constructed using other popular psychometric approaches that do not guarantee unidimensionality, such as reliance on Cronbach alpha (e.g., Dunn, Baguley, & Brunsden, 2014; Green, Lissitz, & Mulaik, 1977) and Principal Component Analysis (e.g., Gorsuch, 1997). Such scales are more likely to incorporate multiple underlying traits, of which only one or some may relate to a particular outcome.

The indicator exclusion procedure provides a statistical method to test whether removal of an indicator causes the trait-outcome correlation to significantly change. This is a very straightforward test of ION, both conceptually and methodologically. The scripts to estimate a p criterion and run the analysis in the R environment are available at <http://www.ut.ee/uku.vainik/ion/>. An even simpler approach is to correlate single items with the outcome, which provides preliminary hints as to which items are related to the trait-outcome relationship.

Regarding limitations, it is important to note that all these methods are based on correlation, and hence are likely to have similar assumptions to data and to the sample size (e.g., Schönbrodt & Perugini, 2013). Further, the effectiveness of ION depends on the sample size and the effect size in question. Hopefully, future optimisations of ION testing can reduce the requirements of sample size. In current Study 2, we were able to test our main association of interest – N5: Impulsiveness-BMI. However, smaller trait-BMI associations remain to be

tested in larger studies. For instance, one could employ the 1958 National Child Development Study that has over 9000 persons (TNS BMRB, 2014). At the same time, larger effect sizes can be well studied in smaller samples (Table 2). The R script published alongside this article can be used to test beforehand, which samples are sufficient for given effect size.

There is good reason to expect that other outcomes than BMI could also have item-specific variance in NEO-PI. As highlighted in introduction, a few recent studies propose that there is considerable meaningful variance left in NEO-PI below facet level (McCrae, 2014; René Mõttus et al., 2015; R. Mõttus et al., 2014). Interpreting item or nuance-level variance might provide similar benefits, as interpreting facets has provided over domains (Briley & Tucker-Drob, 2012; Costa & McCrae, 1995; Judge et al., 2013). The conventional wisdom may be that single items are infused with (random) measurement error and should not be used for substantive research; aggregates may reduce (random) measurement error and therefore better suited for being linked with other variables. This assumption may need a re-assessment. In this study, we observed that often single items predicted BMI better or at least with the same magnitude than the aggregate scores they were part of. In fact, this was true for all of the three facet scales considered in more detail. This may suggest that single items are worth being considered as substantive variables in their own right rather than mere measurement devices.

In conclusion, we have proposed a method that clarifies if trait-outcome associations are caused by scales or particular items. The study illustrates that Spearman's theorem of indifference of indicator is well suited for better understanding personality trait-outcome relationships. We outlined the principles of testing indicator indifference and demonstrated the indicator exclusion procedure in a simulation study. We then applied the procedure to clarify the relationship between personality facets and BMI using a large sample with diverse demographic backgrounds and detailed personality data from self- and informant-reports. We

hope that testing for indifference of indicator will lead to a more precise understanding of personality trait-outcome relationships.

5 Acknowledgements

We would like to thank Delaney Michael Skerret, Kenn Konstabel, Lesley Fellows, Maarika Paaver, Tom Booth, and anonymous reviewers for their valuable comments on earlier versions of the paper, as well as Andres Metspalu for his support.

The Estonian Genome Centre of the University of Tartu was financed by two FP7 grants (201413, 245536). It also received targeted financing from the Estonian Government (SF0180142s08), from the University of Tartu within the framework of the Center of Translational Genomics, and from the European Union through the European Regional Development Fund within the framework of the Centre of Excellence in Genomics. This study was also supported by research funding from the University of Tartu (SP1GVARENG) and by an institutional research funding (IUT2-13) from the Estonian Ministry of Education and Science (IUT2-13). Anu Realo was supported by a grant from the Netherlands Institute for Advanced Study (NIAS) during the preparation of this article.

6 References

- Armon, G., Melamed, S., Shirom, A., Shapira, I., & Berliner, S. (2013). Personality Traits and Body Weight Measures: Concurrent and Across-Time Associations. *European Journal of Personality*, 27(4), 398–408. <http://doi.org/10.1002/per.1902>
- Ashton, M. C., & Lee, K. (2005). Problems with the method of correlated vectors. *Intelligence*, 33(4), 431–444. <http://doi.org/10.1016/j.intell.2004.12.004>

- Bagby, R. M., Vachon, D. D., Bulmash, E. L., Toneatto, T., Quilty, L. C., & Costa, P. T. (2007). Pathological gambling and the five-factor model of personality. *Personality and Individual Differences*, 43(4), 873–880. <http://doi.org/10.1016/j.paid.2007.02.011>
- Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110(2), 305–314. <http://doi.org/10.1037/0033-2909.110.2.305>
- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, 71(3), 425–440. <http://doi.org/10.1007/s11336-006-1447-6>
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110(2), 203–219. <http://doi.org/10.1037/0033-295X.110.2.203>
- Briley, D. A., & Tucker-Drob, E. M. (2012). Broad Bandwidth or High Fidelity? Evidence from the Structure of Genetic and Environmental Effects on the Facets of the Five Factor Model. *Behavior Genetics*, 42(5), 743–763. <http://doi.org/10.1007/s10519-012-9548-8>
- Brummett, B. H., Babyak, M. A., Williams, R. B., Barefoot, J. C., Costa, P. T., & Siegler, I. C. (2006). NEO personality domains and gender predict levels and trends in body mass index over 14 years during midlife. *Journal of Research in Personality*, 40(3), 222–236.
- Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The Convergent Validity between Self and Observer Ratings of Personality: A meta-analytic review. *International Journal of Selection and Assessment*, 15(1), 110–117. <http://doi.org/10.1111/j.1468-2389.2007.00371.x>

- Costa, P. T., Jr, & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment*, 64(1), 21–50. http://doi.org/10.1207/s15327752jpa6401_2
- Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEOFFI) – professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.
- De Fruyt, F., De Bolle, M., McCrae, R. R., Terracciano, A., & Costa, P. T. (2009). Assessing the Universal Structure of Personality in Early Adolescence: The NEO-PI-R and NEO-PI-3 in 24 Cultures. *Assessment*, 16(3), 301–311. <http://doi.org/10.1177/1073191109333760>
- DeYoung, C. G. (2014). Cybernetic Big Five Theory. *Journal of Research in Personality*. <http://doi.org/10.1016/j.jrp.2014.07.004>
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399–412. <http://doi.org/10.1111/bjop.12046>
- Eglit, T., Ringmets, I., & Lember, M. (2013). Obesity, High-Molecular-Weight (HMW) Adiponectin, and Metabolic Risk Factors: Prevalence and Gender-Specific Associations in Estonia. *PLoS ONE*, 8(9), e73273. <http://doi.org/10.1371/journal.pone.0073273>
- Elfhag, K. (2005). Personality correlates of obese eating behaviour: Swedish universities Scales of Personality and the Three Factor Eating Questionnaire. *Eating and Weight Disorders*, 10(4), 210–215.
- Elfhag, K., & Erlanson-Albertsson, C. (2006). Sweet and fat taste preference in obesity have different associations with personality and eating behavior. *Physiology & Behavior*, 88(1–2), 61–66. <http://doi.org/10.1016/j.physbeh.2006.03.006>

- Elfhag, K., & Morey, L. C. (2008). Personality traits and eating behavior in the obese: Poor self-control in emotional and external eating but personality assets in restrained eating. *Eating Behaviors*, 9(3), 285–293. <http://doi.org/10.1016/j.eatbeh.2007.10.003>
- Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The Application of Exploratory Factor Analysis in Applied Psychology: A Critical Review and Analysis. *Personnel Psychology*, 39(2), 291–314. <http://doi.org/10.1111/j.1744-6570.1986.tb00583.x>
- Gerbing, D. W., & Anderson, J. C. (1988). An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment. *Journal of Marketing Research*, 25(2), 186–192. <http://doi.org/10.2307/3172650>
- Gorin, A. A., Phelan, S., Wing, R. R., & Hill, J. O. (2003). Promoting long-term weight control: does dieting consistency matter? *International Journal of Obesity*, 28(2), 278–281. <http://doi.org/10.1038/sj.ijo.0802550>
- Gorsuch, R. L. (1997). Exploratory Factor Analysis: Its Role in Item Analysis. *Journal of Personality Assessment*, 68(3), 532–560. http://doi.org/10.1207/s15327752jpa6803_5
- Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of Coefficient Alpha as an Index of Test Unidimensionality1. *Educational and Psychological Measurement*, 37(4), 827–838. <http://doi.org/10.1177/001316447703700403>
- Horn, J. L., & McArdle, J. J. (2007). Understanding human intelligence since Spearman. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 205–247). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Houben, K., Nederkoorn, C., & Jansen, A. (2013). Eating on impulse: The relation between overweight and food-specific inhibitory control. *Obesity*, n/a–n/a. <http://doi.org/10.1002/oby.20670>

- Iacovino, J. M., Powers, A. D., & Oltmanns, T. F. (2014). Impulsivity mediates the association between borderline personality pathology and body mass index. *Personality and Individual Differences*, 56, 100–104.
<http://doi.org/10.1016/j.paid.2013.08.028>
- Jen, A., Saunders, E. F., Ornstein, R. M., Kamali, M., & McInnis, M. G. (2013). Impulsivity, anxiety, and alcohol misuse in bipolar disorder comorbid with eating disorders. *International Journal of Bipolar Disorders*, 1(1), 1–9.
- Jensen, A. R. (1998). *The g factor: the science of mental ability*. Westport, Conn: Praeger.
- Johnson, W., Bouchard Jr., T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: consistent results from three test batteries. *Intelligence*, 32(1), 95–107.
[http://doi.org/10.1016/S0160-2896\(03\)00062-X](http://doi.org/10.1016/S0160-2896(03)00062-X)
- Johnson, W., Nijenhuis, J. te, & Bouchard, T. J. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence*, 36(1), 81–95.
<http://doi.org/10.1016/j.intell.2007.06.001>
- Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: integrating three organizing frameworks with two theoretical perspectives. *The Journal of Applied Psychology*, 98(6), 875–925.
<http://doi.org/10.1037/a0033901>
- Kallasmaa, T., Allik, J., Realo, A., & McCrae, R. R. (2000). The Estonian version of the NEO-PI-R: An examination of universal and culture-specific aspects of the five-factor model. *European Journal of Personality*, 14(3), 265–278. [http://doi.org/10.1002/1099-0984\(200005/06\)14:3<265::AID-PER376>3.0.CO;2-B](http://doi.org/10.1002/1099-0984(200005/06)14:3<265::AID-PER376>3.0.CO;2-B)

- Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the Accuracy of Personality Judgments by the Self and Knowledgeable Others. *Journal of Personality*, 64(2), 311–337. <http://doi.org/10.1111/j.1467-6494.1996.tb00513.x>
- Leitsalu, L., Haller, T., Esko, T., Tammesoo, M.-L., Alavere, H., Snieder, H., ... Metspalu, A. (2014). Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *International Journal of Epidemiology*, dyt268. <http://doi.org/10.1093/ije/dyt268>
- Lubke, G. H., Dolan, C. V., & Kelderman, H. (2001). Investigating group differences on cognitive tests using Spearman's hypothesis: An evaluation of Jensen's method. *Multivariate Behavioral Research*, 36(3), 299–324.
- Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, 70(3), 151–159.
- McCrae, R. R. (2014). A More Nuanced View of Reliability Specificity in the Trait Hierarchy. *Personality and Social Psychology Review*, 1088868314541857. <http://doi.org/10.1177/1088868314541857>
- McCrae, R. R., & Costa, P. T., Jr. (2010). *NEO Inventories professional manual*. Odessa, FL: Psychological Assessment Resources.
- Mõttus, R., Luciano, M., Starr, J. M., Pollard, M. C., & Deary, I. J. (2013). Personality Traits and Inflammation in Men and Women in Their Early 70s: The Lothian Birth Cohort 1936 Study of Healthy Aging. *Psychosomatic Medicine January 2013*, 75(1), 11–19. <http://doi.org/10.1097/PSY.0b013e31827576cc>
- Mõttus, R., McCrae, R. R., Allik, J., & Realo, A. (2014). Cross-rater agreement on common and specific variance of personality scales and items. *Journal of Research in Personality*, 52, 47–54. <http://doi.org/10.1016/j.jrp.2014.07.005>

- Mõttus, R., Realo, A., Allik, J., Esko, T., Metspalu, A., & Johnson, W. (2015). Within-Trait Heterogeneity in Age Group Differences in Personality Domains and Facets: Implications for the Development and Coherence of Personality Traits. *PLoS ONE*, *10*(3), e0119667. <http://doi.org/10.1371/journal.pone.0119667>
- Ogden, C. L., Carroll, M. D., Curtin, L. R., McDowell, M. A., Tabak, C. J., & Flegal, K. M. (2006). Prevalence of overweight and obesity in the United States, 1999-2004. *JAMA: The Journal of the American Medical Association*, *295*(13), 1549–1555. <http://doi.org/10.1001/jama.295.13.1549>
- Ostendorf, F., & Angleitner, A. (1994). A comparison of different instruments proposed to measure the Big Five. *European Review of Applied Psychology / Revue Européenne de Psychologie Appliquée*, *44*(1), 45–53.
- Oswald, L. M., Wong, D. F., Zhou, Y., Kumar, A., Brasic, J., Alexander, M., ... Wand, G. S. (2007). Impulsivity and chronic stress are associated with amphetamine-induced striatal dopamine release. *NeuroImage*, *36*(1), 153–166. <http://doi.org/10.1016/j.neuroimage.2007.01.055>
- Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the Prediction of Consequential Outcomes. *Annual Review of Psychology*, *57*(1), 401–421. <http://doi.org/10.1146/annurev.psych.57.102904.190127>
- Pachucki, M. A. (2012). Food pattern analysis over time: unhealthy eating trajectories predict obesity. *International Journal of Obesity*, *36*(5), 686–694. <http://doi.org/10.1038/ijo.2011.133>
- Rasmussen, E. B., Lawyer, S. R., & Reilly, W. (2010). Percent body fat is related to delay and probability discounting for food in humans. *Behavioural Processes*, *83*, 23–30. <http://doi.org/10.1016/j.beproc.2009.09.001>

- R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Revelle, W. (2014). psych: Procedures for personality and psychological research. *Northwestern University, Evanston. R Package Version, 1(1)*.
- Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes. *Perspectives on Psychological Science, (4)*, 313.
- Rolls, B. J., Fedoroff, I. C., & Guthrie, J. F. (1991). Gender differences in eating behavior and body weight regulation. *Health Psychology, 10(2)*, 133–142.
<http://doi.org/10.1037/0278-6133.10.2.133>
- Ruiz, M. A., Pincus, A. L., & Dickinson, K. A. (2003). NEO PI-R Predictors of Alcohol Use and Alcohol-Related Problems. *Journal of Personality Assessment, 81(3)*, 226–236.
http://doi.org/10.1207/S15327752JPA8103_05
- Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47(5)*, 609–612.
<http://doi.org/10.1016/j.jrp.2013.05.009>
- Spearman, C. (1927). *The Abilities Of Man*. Macmillan And Company., Limited. Retrieved from <http://archive.org/details/abilitiesofman031969mbp>
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87(2)*, 245.
- Sutin, A. R., Costa, P. T., Jr, Chan, W., Milaneschi, Y., Eaton, W. W., Zonderman, A. B., ... Terracciano, A. (2013). I Know Not To, but I Can't Help It: Weight Gain and Changes in Impulsivity-Related Personality Traits. *Psychological Science, 24(7)*, 1323–1328.
<http://doi.org/10.1177/0956797612469212>

- Sutin, A. R., Evans, M. K., & Zonderman, A. B. (2013). Personality traits and illicit substances: The moderating role of poverty. *Drug and Alcohol Dependence*, 131(3), 247–251. <http://doi.org/10.1016/j.drugalcdep.2012.10.020>
- Sutin, A. R., Ferrucci, L., Zonderman, A. B., & Terracciano, A. (2011). Personality and obesity across the adult life span. *Journal of Personality and Social Psychology*, 101(3), 579–592.
- Sutin, A. R., Milaneschi, Y., Cannas, A., Ferrucci, L., Uda, M., Schlessinger, D., ... Terracciano, A. (2012). Impulsivity-related Traits Are Associated with Higher White Blood Cell Counts. *Journal of Behavioral Medicine*, 35(6), 616–623.
<http://doi.org/10.1007/s10865-011-9390-0>
- Sutin, A. R., Rogers, D. L., Mercado, A., Weimer, A., Rodriguez, C. C., Gonzalez, M., ... Terracciano, A. (2015). The association between personality traits and body mass index varies with nativity among individuals of Mexican origin. *Appetite*.
<http://doi.org/10.1016/j.appet.2015.02.036>
- Sutin, A. R., Zonderman, A. B., Uda, M., Deiana, B., Taub, D. D., Longo, D. L., ... Terracciano, A. (2013). Personality Traits and Leptin. *Psychosomatic Medicine*, 75(5), 505–509. <http://doi.org/10.1097/PSY.0b013e3182919ff4>
- Tekkel, M., & Veideman, T. (2013). *Eesti täiskasvanud rahvastiku tervisekäitumise uuring, 2012 [Health Behavior among Estonian Adult Population, 2012]*. Tallinn: NIHD.
Retrieved from
https://intra.tai.ee/images/prints/documents/136479842690_TKU_2012.pdf
- Terracciano, A., Sutin, A. R., McCrae, R. R., Deiana, B., Ferrucci, L., Schlessinger, D., ... Costa, P. T. (2009). Facets of personality linked to underweight and overweight. *Psychosomatic Medicine*, 71(6), 682–689.
<http://doi.org/10.1097/PSY.0b013e3181a2925b>

- TNS BMRB. (2014). Technical Report of the 1958 National Child Development Study: Age 55 survey (2013/2014). Retrieved from <http://www.cls.ioe.ac.uk/shared/get-file.ashx?id=1942&itemtype=document>
- Tsukayama, E., Duckworth, A. L., & Kim, B. (2012). Resisting everything except temptation: Evidence and an explanation for domain-specific impulsivity. *European Journal of Personality*, 26(3), 318–334. <http://doi.org/10.1002/per.841>
- Vainik, U., Dagher, A., Dubé, L., & Fellows, L. K. (2013). Neurobehavioural correlates of body mass index and eating behaviours in adults: A systematic review. *Neuroscience & Biobehavioral Reviews*, 37(3), 279–299. <http://doi.org/10.1016/j.neubiorev.2012.11.008>
- Vainik, U., Dubé, L., Lu, J., & Fellows, L. K. (2015). Personality and situation predictors of consistent eating patterns. *Submitted for Publication*.
- Vainik, U., Neseliler, S., Konstel, K., Fellows, L. K., & Dagher, A. (2015). Eating traits questionnaires as a continuum of a single concept. Uncontrolled eating. *Appetite*, 90, 229–239. <http://doi.org/10.1016/j.appet.2015.03.004>
- Villafuerte, S., Heitzeg, M. M., Foley, S., Wendy Yau, W.-Y., Majczenko, K., Zubieta, J.-K., ... Burmeister, M. (2012). Impulsiveness and insula activation during reward anticipation are associated with genetic variants in GABRA2 in a family sample enriched for alcoholism. *Molecular Psychiatry*, 17(5), 511–519. <http://doi.org/10.1038/mp.2011.33>
- Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., Lumley, T., ... Venables, B. (2014). gplots: Various R programming tools for plotting data (Version 2.14.2). Retrieved from <http://cran.r-project.org/web/packages/gplots/index.html>
- Wing, R. R., & Phelan, S. (2005). Long-term weight loss maintenance. *The American Journal of Clinical Nutrition*, 82(1), 222S–225S.

Table 1. Percent of correctly excluded items across scenarios lacking ION

β	Scenario	Sample size						
		100	250	500	750	1,000	2,500	5,000
0.35	7+1	46.02	76.15	93.74	98.31	99.42	100	100
	6+2	61.08	86.15	97.12	99.41	99.89	100	100
	6+1	38.65	67.82	90.19	97.28	99.17	100	100
	5+3	67.83	86.82	96.31	99	99.75	100	100
	5+2	56.79	81.46	95.86	98.96	99.77	100	100
	5+1	37.13	62.82	87.84	96.08	98.74	99.98	100
0.25	7+1	30.65	53.44	77.8	89.09	94.67	99.94	100
	6+2	46.18	68.96	87.39	95.02	97.91	99.97	100
	6+1	28.69	47.29	70.38	84.29	91.89	99.84	100
	5+3	55.4	73.61	88.42	94.02	97.06	99.89	100
	5+2	44.81	64.58	82.92	92.5	96.08	99.96	100
	5+1	27.56	44	66.18	80.84	89.66	99.66	100
0.15	7+1	19.68	31.3	45.89	57.61	67.77	93.43	99.35
	6+2	32.72	45.23	61.05	72.94	80.25	97.21	99.87
	6+1	19.2	27.45	39.6	50.04	60.06	90.5	99
	5+3	44.71	54.22	67.7	75.68	81.75	96.6	99.65

5+2	34.3	42.8	57.29	66.82	74.54	95.65	99.69
5+1	20.37	26.21	36.82	47.4	56.87	87.56	98.61

Note. 10,000 simulations per cell. β = association strength between trait (T_2) and outcome.

See text for detailed explanation of scenarios.

Table 2. Optimal p -values for different trait-outcome associations and sample sizes

β	Sample size						
	100	250	500	750	1000	2500	5000
0.35	-	-	0.039	0.018	8.6×10^{-3}	1.7×10^{-4}	1.2×10^{-7}
0.25	-	-	-	0.055	0.036	3.7×10^{-3}	9.0×10^{-5}
0.15	-	-	-		-	0.041	9.4×10^{-3}

Note: β = association between trait and outcome; - = optimal p -value cannot be determined

Table 3. Residualised correlation coefficients between personality traits and body mass index (BMI), across self- and informant-ratings.

Domain / Facet	Self	Informant
	r_p	r_p
Neuroticism	0.03	0.04
Extraversion	0.04	0.03
Openness	0.00	0.02
Agreeableness	0.00	-0.01
Conscientiousness	-0.04	-0.05
N1: Anxiety	0.01	0.03
N2: Angry Hostility	0.02	-0.01
N3: Depression	0.01	0.03
N4: Self-consciousness	0.03	0.01
<u>N5: Impulsiveness</u>	0.13	0.15
N6: Vulnerability	-0.04	-0.02
E1: Warmth	0.04	0.06
E2: Gregariousness	0.00	0.02
<u>E3: Assertiveness</u>	0.07	0.06
E4: Activity	-0.01	-0.02

E5: Excitement-Seeking	0.00	0.00
E6: Positive Emotions	0.07	0.04
O1: Fantasy	0.04	0.03
O2: Aesthetics	0.01	-0.01
O3: Feelings	0.00	0.03
O4: Actions	-0.02	0.01
O5: Ideas	0.00	-0.01
O6: Values	-0.03	0.01
A1: Trust	-0.02	-0.01
A2: Straightforwardness	-0.03	-0.04
A3: Altruism	0.01	0.02
A4: Compliance	-0.03	0.01
A5: Modesty	0.02	0.00
A6: Tender-mindedness	0.04	-0.02
C1: Competence	0.02	0.00
<u>C2: Order</u>	-0.06	-0.07
C3: Dutifulness	-0.03	-0.02
C4: Achievement Striving	-0.04	-0.05

C5: Self-Discipline	-0.02	-0.06
C6: Deliberation	-0.02	-0.01

Note: Body mass index (BMI), has been residualised for the effects of age, gender, and education. Bold = significant at $p < 0.01$. r_p = residualised correlation coefficient. Underlined facets reported effects in both self- and informant-ratings and were used in further analysis.

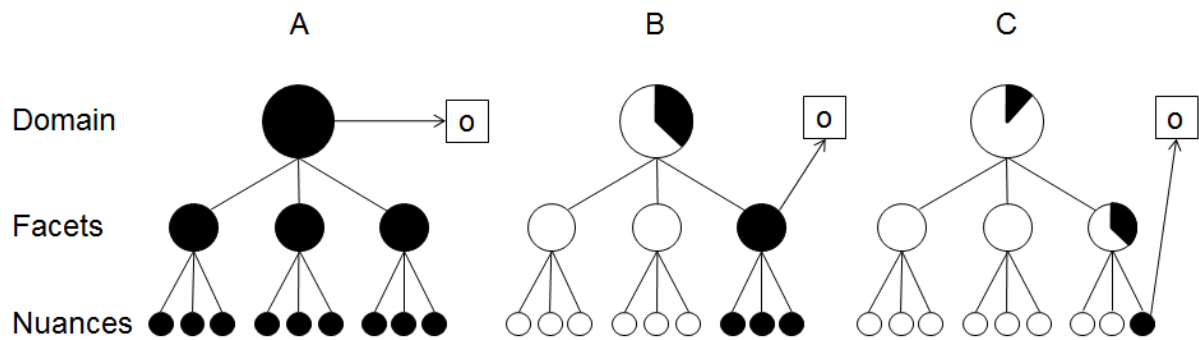


Figure 1. Trait-outcome associations at different levels of personality measurement. Filled circle = trait's association with an outcome that is indicator indifferent. Semi-filled circle = trait's association with an outcome that depends on a certain indicator. Empty circle = no association between trait and outcome. A. Outcome relates to the domain-level trait. Therefore, the outcome also relates to all domain's indicators (i.e. facets), and also to all indicators of facets (i.e. items, also interpretable as nuances (McCrae, 2014)). B. Outcome relates to the facet-level trait. The outcome could also correlate with the domain-level trait sum-score. However, because the outcome only correlates with a single indicator of the domain, the outcome pertains only to the facet level as such. C. Outcome relates to a single nuance. The association might be visible also at domain and facet level sum-scores. However, in both cases no other indicators relate to the outcome. Therefore, the association pertains to the nuance level.

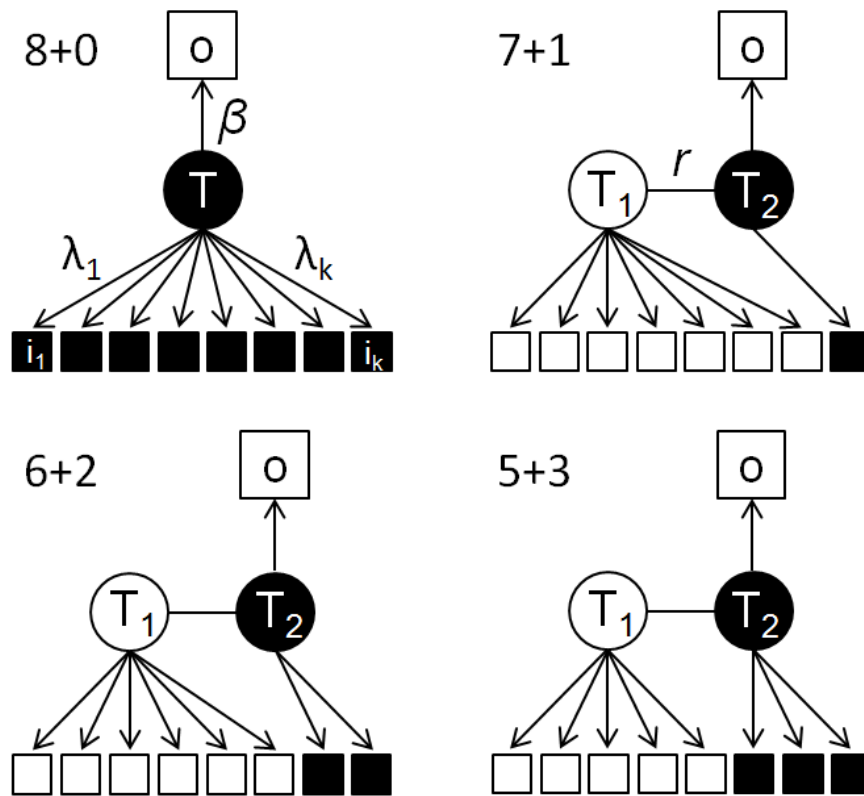


Figure 2. Diagrams of the different scales modelled. In 8+0 scenario, the scale-outcome association is indicator indifferent. In other scenarios, the scale-outcome association pertains to a subset of items reflecting a separate trait (T_2). Rectangles are observed items or outcomes, circles denote latent variables. Shapes filled with black denote outcome-related traits and indicators, white shapes are not related to an outcome. Model parameters are shown once, although they apply in several scenarios. β = trait-outcome association; λ = factor loading; i = item; k = number of items in a scale, here $k = 8$; o = outcome; r = correlation between traits; T = trait; T_1 = trait not related to outcome; T_2 = trait related to outcome.

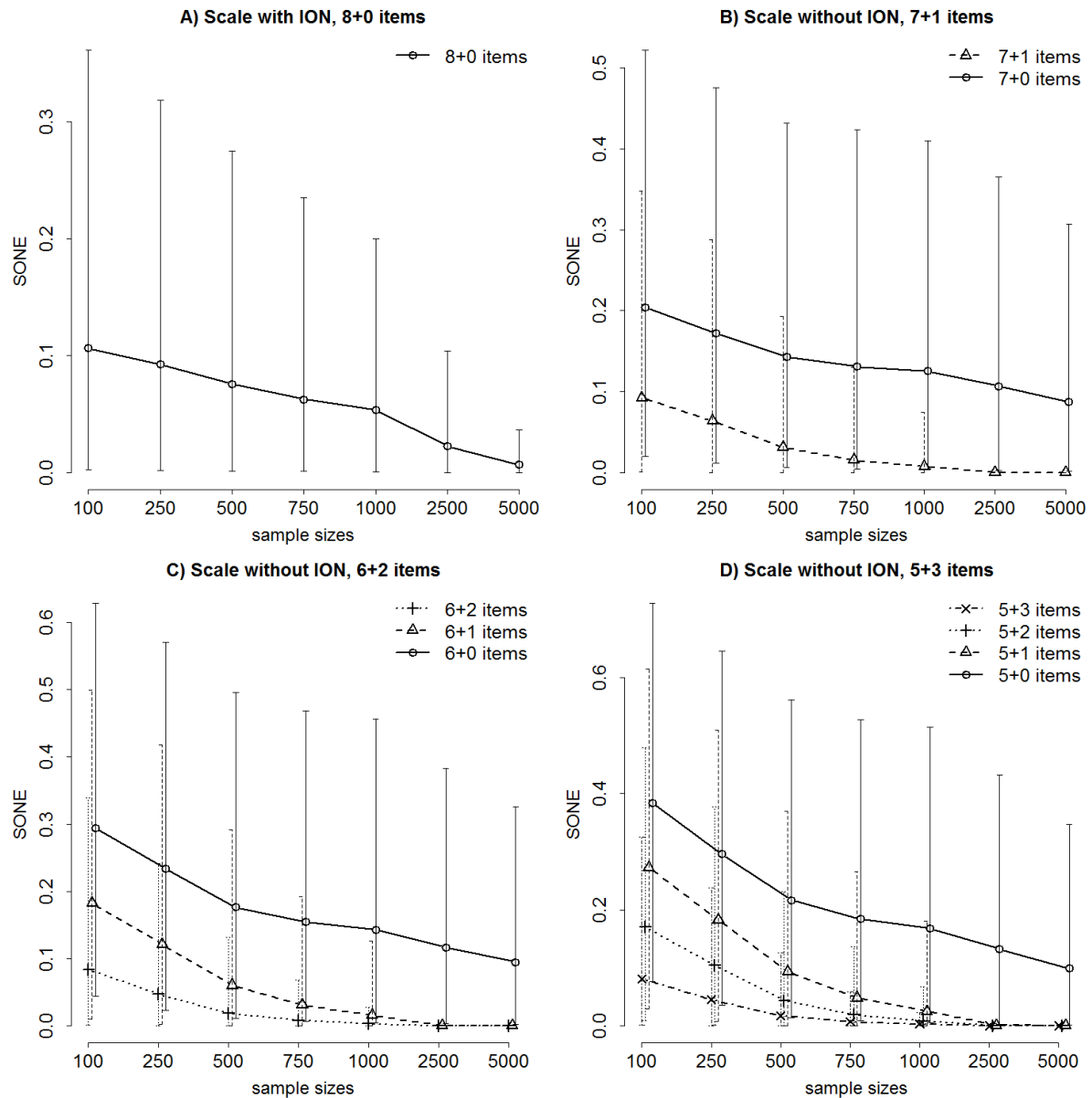


Figure 3. Mean minimum “significance of indicator exclusion” (SONE) values across different simulation conditions and different sample sizes. A) The scale was designed to have an indicator indifferent (ION) relationship with an outcome. B-D) The scale-outcome relationship was not indicator indifferent, such that most items correlated with the main trait (T_1), but the outcome related to a sub-trait (T_2) represented by 1-3 items only. The T_2 -related items were then iteratively removed (e.g., 5 + 2 refers to a scale from which one item had been removed). Trait-outcome association (β) = 0.25

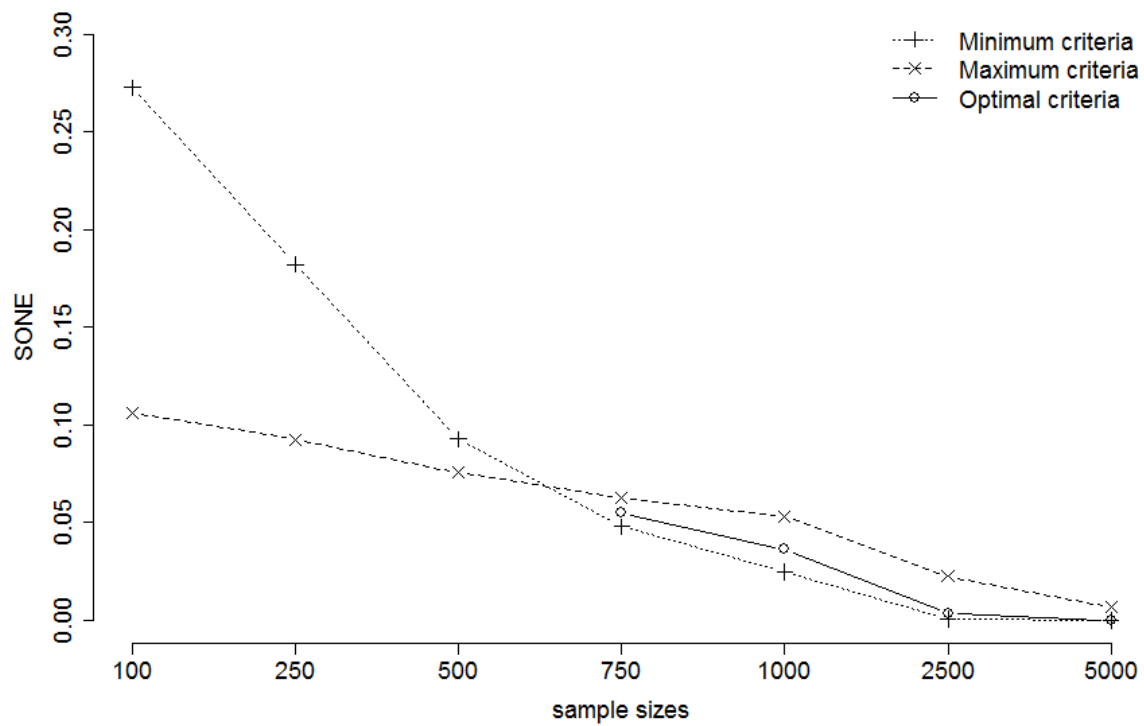


Figure 4. Detecting optimal p criteria. Minimum criteria: highest “significance of indicator exclusion” (SONE) values that excluded items causing lack of indifference of indicator (scenario 5+1 from Figure 3D). Maximum criteria: lowest SONE values that did not exclude items that belonged to a trait related to outcome (scenario 8+0 from Figure 3A). Optimal criteria: geometric mean between minimum and maximum. Trait-outcome association (β) = 0.25

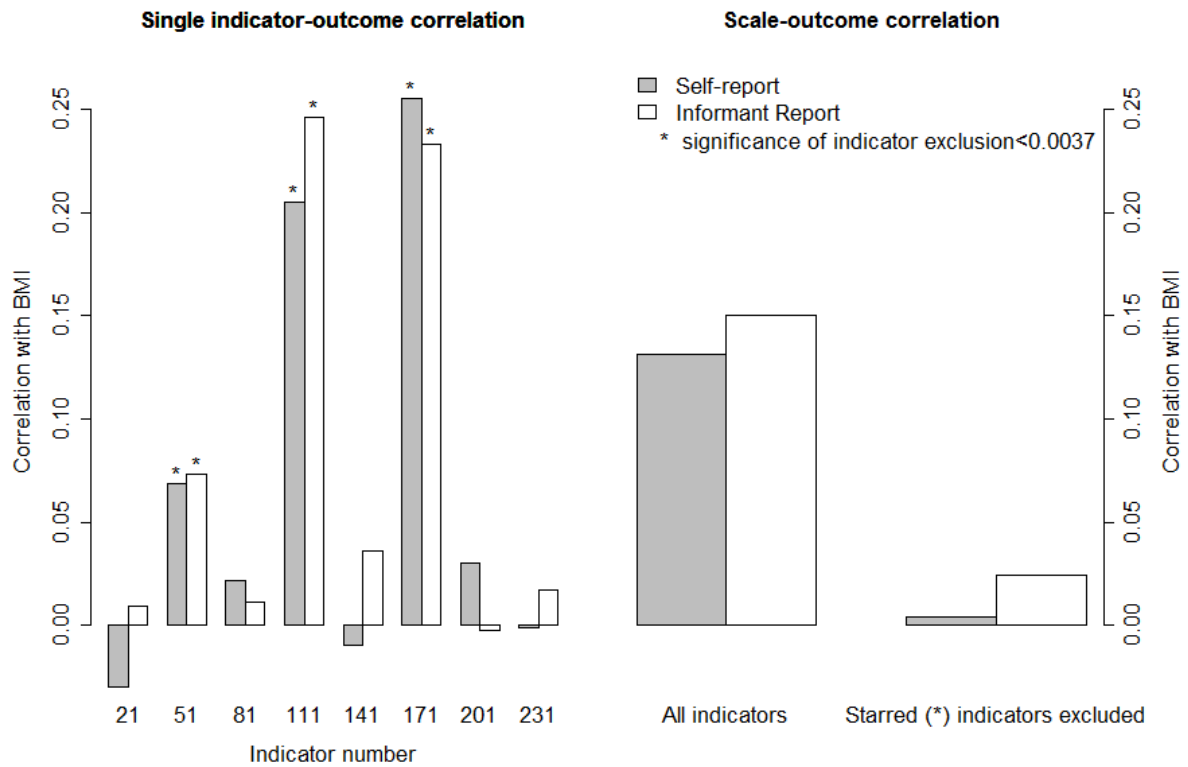


Figure 5. Testing relationship between body mass index (BMI), and N5: Impulsiveness for indifference of the indicator with indicator exclusion procedure.

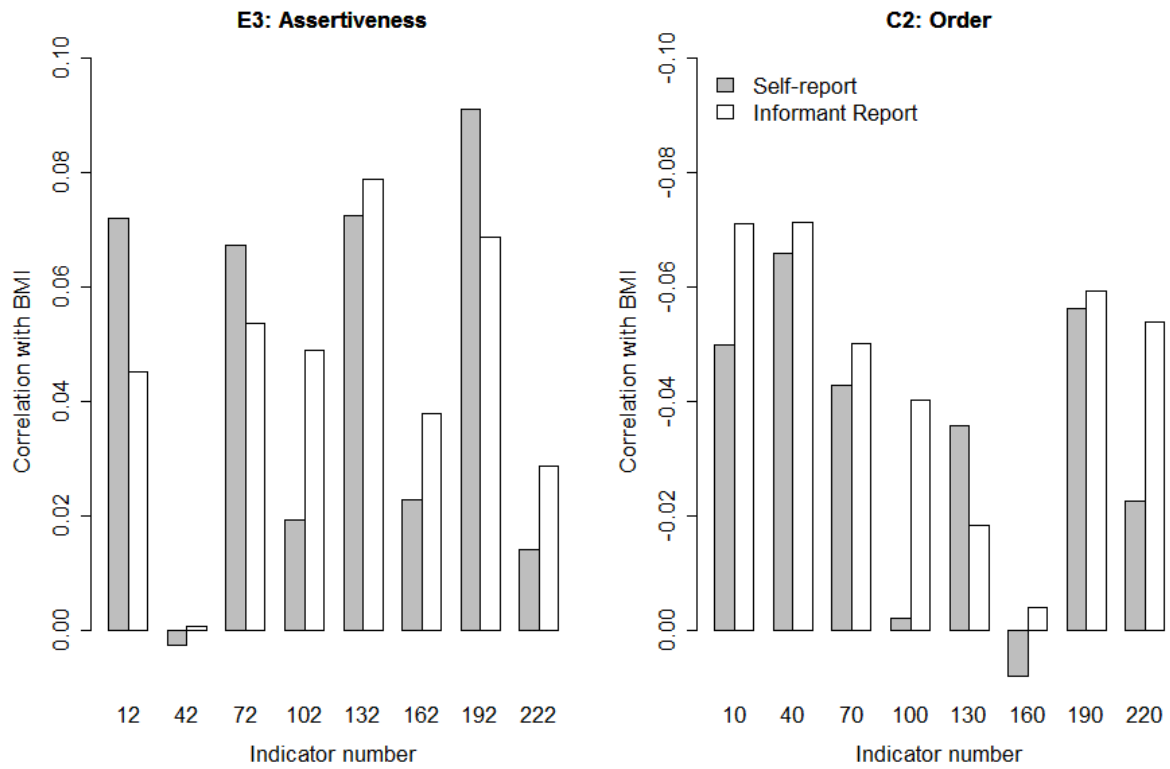


Figure 6. Single item-outcome associations between body mass index (BMI), E3: Assertiveness and C2: Order